ERIC Notebook

May/June 2001 Issue 18

UNC SCHOOL OF PUBLIC HEALTH DEPARTMENT OF EPIDEMIOLOGY

Michel Ibrahim, MD, PhD
Director of Education Program

Lorraine Alexander, DrPH

Carl Shy, MD, DrPH

Sandra Deming, MPH
Graduate Research Asst.

VA ERIC EPIDEMIOLOGIC RESEARCH & INFORMATION CENTER Durham, North Carolina

Ron Horner, PhD
ERIC Director

Beth Armstrong
ERIC Program Manager

http://hsrd.durham.med.va.gov/ERIC/

The ERIC Notebook is funded by the Department of Veterans Affairs (DVA), Veterans Health Administration (VHA), Cooperative Studies Program (CSP), to promote the strategic growth of the epidemiologic capacity of the DVA.

Common Statistical Tests and Applications in Epidemiological Literature

Any individual in the medical field will, at some point, encounter instances when epidemiological methods and statistics will be valuable tools in addressing research questions of interest.

Examples of such questions might include:

- Will treatment with a new anti-hypertensive drug significantly lower mean systolic blood pressure?
- Is a visit with a social worker, in addition to regular medical visits, associated with greater satisfaction of care for cancer patients as compared to those who only have regular medical visits?

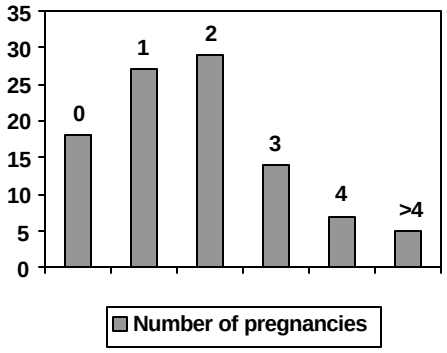
There are a number of steps in evaluating data before actually addressing the above questions. These steps include description of your data, as well as determining what are the appropriate tests for your data.

Description of Data

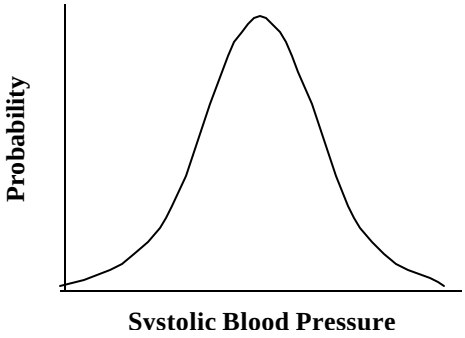
The type of data one has determines the statistical procedures that are utilized. Data is typically described in a number of ways: by type, distribution, location and variation.

There are three different types of data: *nominal*, *ordinal*, and *continuous* data. *Nominal* data does not have an established order or rank, and contains a finite number of values. Gender and race are examples of nominal data. *Ordinal* data has a limited number of values between which no other possible values exist. Number of children and stage of disease are good examples of ordinal data. It should be noted that ordinal data does not have to have evenly spaced values as occurs with continuous data, however, there is an implied underlying order. Since both ordinal and nominal data have a finite number of possible values, they are referred to as *discrete* data as well. The last type of data is *continuous* data which is characterized by having an infinite number of evenly spaced values. Blood pressure and age fall into this category. It should be noted for data collection and analysis that continuous, ordinal, or nominal values can be grouped. Grouped data are often referred to as *categorical.* Possible categories might include: low, medium, high, or those representing a numerical range.

A second characteristic of data description, distribution, refers to the frequencies or probabilities with which values occur within our population. Discrete data is often represented graphically with bar graphs like the one below:
Continuous data is typically represented by a



frequency polygon seen below: This symmetric, bell-shaped curve is known as a Gaussian distribution, the most commonly assumed distribution in statistical analysis.



Hypothesis Testing

Hypothesis testing, also known as statistical inference or significance testing, involves testing a specified hypothesized condition for a population's parameter. This condition is best described as the null hypothesis. For example, in a clinical trial of a new anti-hypertensive drug, the null hypothesis would state that there is no difference in effect when comparing the new drug to the current standard treatment. Contrary to the null is the alternative hypothesis which generally defines the possible values for a parameter of interest. For the previous example, the alternative hypothesis is that there is a difference in the mean blood pressure of the standard treatment and new drug group following therapy. The alternative hypothesis might also be described as your "best guess" as to what the values are.

However, in statistical analysis, the null hypothesis is the main interest, and is the one actually being tested. In statistical testing, we assume that the null hypothesis is correct, and determine how likely we are to have obtained the sample (or values) we actually obtained in our study under the condition of the null. If we determine that the probability of obtaining the sample we observed, is sufficiently small, then we can reject the null hypothesis. Since we are able to reject the null hypothesis, we can therefore conclude that the alternative hypothesis is true.

On the other hand, if the probability of obtaining our study results is not small, we *fail to reject* the assumption that the null hypothesis is true. It should be noted that we are not concluding that the null is true. This is a small, but important distinction. A test that fails to reject the null hypothesis should be considered inconclusive. An example will help to illustrate this point.

In a sealed bag, we have 100 blue marbles and 20 red marbles. (This bag is essentially representing the entire population). One individual formulates the null hypothesis that "all the marbles are blue", and the alternative which is "all the marbles are not blue". To test this hypothesis, 10 marbles are sampled from the bag. All ten marbles selected are indeed blue. Thus he has failed to reject the null that all the marbles in the bag are blue. However, because all of the marbles were not sampled, you cannot conclude that all the marbles in the bag are blue. (We happen to know this is not true, but it is impossible to know in the real world with populations too large to fully evaluate). If another individual selects 10 marbles from the bag, and finds that 8 are blue and 2 are red, we can reject the null hypothesis that all the marbles are blue since we have selected at least one red marble.

Error in statistical testing:

Earlier, we indicated that we can reject the null hypothesis if the probability of obtaining a sample like the one observed in our study is sufficiently small. You may ask "What is sufficiently small?" How small is determined by how willing we are to reject the null hypothesis when it accurately reflects the population from which it is sampled. This type of error is called a *Type I error*. This error is also commonly called alpha (α). α is the probability of rejecting the null hypothesis when the null is true. This probability is selected by the researcher and is typically set at 0.05. It is important to remember that this is an arbitrary cut-point, and should be taken into consideration when making conclusions about the results of the study (See previous ERIC notebook for more details).

There is a second type of error that can be made during statistical testing. It is known as *Type II error* which is the probability of not rejecting the null when the alternative hypothesis is indeed true, or in other words, accepting the null when the null hypothesis is false. *Type II error* is commonly know as β. β relates to another important parameter in statistical testing which is *power*. Power is equal to $(1-\beta)$, and is essentially the ability to avoid making a type II error. Like α, power is also defined by the researcher, and is typically set at 0.80. Below is a schematic of the relationships between α, β and power.

	Truth	
Decision	**Null True**	**Null False**
Reject Null	α	power
Accept Null		β

Students' T test

This test is most commonly used to test the difference between the means of the dependent variables of two groups. For example, this test would be appropriate if one wanted to evaluate whether or not a new anti-hypertensive drug reduces mean systolic blood pressure.

Example:

To evaluate if drug Z reduces mean systolic blood pressure, a randomized clinical trial will be performed where 12 individuals receive drug Z and 8 receive a placebo. The null hypothesis to be tested is that there is no difference in the mean systolic blood pressure of the experimental and placebo groups. The alternative hypothesis is that there is a difference between the means of the two groups. The type I error for your trial will be 5%.

Results:

Below is the group assignments and resulting SBP

Patient	Assignment	Systolic BP
1	Drug Z	100
3	Drug Z	110
5	Drug Z	122
7	Drug Z	109
9	Drug Z	108
11	Drug Z	111
13	Drug Z	118
15	Drug Z	105
17	Drug Z	115
18	Drug Z	119
19	Drug Z	106
20	Drug Z	109
2	Placebo	129
4	Placebo	125
6	Placebo	136
8	Placebo	129
10	Placebo	135
12	Placebo	134
14	Placebo	140
16	Placebo	128

$$\text{mean}_{drug} = \frac{100 + 110 + \ldots + 109}{12} = 111 \text{ mm Hg}$$

$$\text{mean}_{placebo} = \frac{129 + 125 + \ldots + 128}{8} = 132 \text{ mm Hg}$$

$$\text{mean}_{drug} - \text{mean}_{placebo} = -21 \text{ mm Hg}$$

Now that we have determined the difference between means, we need to determine the standard error for that difference which calculated using the pooled estimate of the variance (s).

The formula for the standard error of the drug Z group is:

$$s^2_{drug} = \frac{\Sigma (SBP_{drug} - \text{mean}_{drug})^2}{n_{drug} - 1} =$$

$$s^2_{drug} = \frac{[(100-111)^2 + (110-111)^2 + \ldots + (109-111)^2]}{12-1} = 40.9$$

$$s^2_{placebo} = 25.1$$

The standard error for the placebo group is calculated in the same manner substituting the values for the placebo group.

Next, we would need to calculate a pooled estimate of the variance using the following equation:

$$s^2_p = \frac{[(n_{drug} - 1)\, s^2_{drug}] + [(n_{placebo} - 1)\, s^2_{placebo}]}{(n_{drug} - 1) + (n_{placebo} - 1)} =$$

$$s^2_p = \frac{(11)(40.9) + (7)(25.1)}{11 + 7} = \frac{626}{18} = 34.8$$

The pooled estimate of the variance can then be utilized to calculate the standard error for the difference in means:

$$SE^2 (\text{mean}_{drug} - \text{mean}_{placebo}) = \frac{s^2_p}{n_{drug}} + \frac{s^2_p}{n_{placebo}}$$

$$SE^2 = \frac{34.8}{12} + \frac{34.8}{8} = 7.236$$

$$SE = 2.69$$

Now we are finally ready to test for significant differences in the mean blood pressure of our two groups: (*mean indicates the hypothesized values for the null-generally this quantity would = 0 when there is no difference expected between the drug and placebo groups).

$$t = \frac{(\text{mean}_{drug} - \text{mean}_{placebo}) - (*\text{mean}_{drug} - *\text{mean}_{placebo})}{SE (\text{mean}_{drug} - \text{mean}_{placebo})}$$

$$t = \frac{-21 - 0}{2.69} = -7.8 = |-7.8| = 7.8$$

We now compare our calculated value to a table of critical values for the Students' T distribution (found in most basic statistics books). The table also requires that we know the *degrees of freedom* and the value of α we have selected. Degrees of freedom (df) refers to the amount of information that a sample has in estimating the variance. It is generally the sample size minus one. The df for our calculation is 12 + 8 - 2 = 18 (the sample size for each group - 1). With a two tailed α of 0.05, our value |7.8| is greater than the critical value from the table (2.101). Thus, we can reject the null hypothesis that there is no difference between mean blood pressure levels, and accept, by elimination, our alternative hypothesis.

Chi-Square Analysis:

What happens if we don't have continuous data, and are faced with categorical data instead? We could turn to chi-square analysis to evaluate if there are significant associations between a given exposure and outcome (the row and column variables in a contingency table). 2 X 2 contingency tables are one of the most common ways to present categorical data, and we can see this in analyzing data that was collected to address the question presented in Issue 17:

Is a visit with a social worker, in addition to regular medical visits, associated with greater satisfaction of care for cancer patients as compared to those who only have regular medical visits?

Below is a generic 2 X 2 table representing the data. It is important to note the set-up of the table as cell *a* generally represents the group of interest (diseased and exposed) and cell *d* represents the referent group (no disease and unexposed).

Column Value	**Row value** 1	2	**Total**
1	*a*	*b*	*a + b*
2	*c*	*d*	*c + d*
Total	*a + c*	*b + d*	*n*

Here we have the contingency table with data from our trial:

Social Worker Visit?	**Greater Satisfaction?** Yes	No	**Total**
Yes	64	46	90
No	36	54	110
Total	100	100	200

In chi-square analysis we are testing the null hypothesis that there is no association between a social worker visit and with a greater satisfaction with care.

Generally, in evaluating this type of data, it is important for each of the individual cells to have large values, (i.e greater than 5 or 10 each), If these conditions are not met, a special type of chi-square analysis is conducted called the EXACT test. This will not be discussed in this notebook.

To calculate the chi-square statistic ($\div^2$):

$$\div^2 = \sum \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i}$$

with *i* representing the frequency in a particular cell of the 2 X 2 table. Below is the calculation for the frequencies that are **expected** in each cell.

Column Value	**Row value** 1	2	**Total**
1	$\frac{(a+b)(a+c)}{n}$	$\frac{(a+b)(b+d)}{n}$	*a + b*
2	$\frac{(c+d)(a+c)}{n}$	$\frac{(c+d)(b+c)}{n}$	*c + d*
Total	*a + c*	*b + d*	*n*

Thus, we now have a table that has both the actual and expected (in parentheses):

Social Worker Visit?	**Greater Satisfaction?** Yes	No	**Total**
Yes	64 (55)	46 (55)	90
No	36 (45)	54 (45)	110
Total	100	100	200

With this information, we can now calculate the $\div^2$ statistic:

$$\div^2 = \sum \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i}$$

$$\div^2 = \frac{(64-55)^2}{55} + \frac{(46-55)^2}{55} + \frac{(36-45)^2}{45} + \frac{(54-45)^2}{45}$$

$$\div^2 = 6.545$$

The chi-square statistic for this data has approximately 1 degree of freedom, an α of 0.05 and it is compared to the critical values on standard Chi-square table. Note that the degrees of freedom would increase as the number of rows and columns of our tables increases (for instance a 3 X 4 table). Since our calculated value ($\div^2$ = 6.545) is greater than the critical value (3.841), we can once again reject the null hypothesis that there is no association between the exposure and the outcome of interest, and conclude that in this case seeing a social worker is significantly associated with a greater satisfaction with care.

Important notes:

It is important to remember that the statistical tests and examples presented here are only an elementary presentation of the large scope of situations that can be addressed by this data. The intention of this notebook is to provide a basic understanding of the underlying principles of these statistical tests rather than implying that what has been presented is appropriate for every situation.

Further information about these statistical tests, and other applications can be found in the following references:

Statistical First Aid: Interpretation of Health Research Data by Robert P Hirsch and Richard K. Riegelman. Blackwell Scientific Publications, Cambridge, MA 1992.

Categorical Data Analysis, Using the SAS System by ME Stokes, CS Davis, and GG Koch. SAS Institute Inc., Cary, NC, 1995.

Self Evaluation:

Q1: **True or False:** In hypothesis testing, we are testing the alternative hypothesis.

Q2: **Select the incorrect classification of data:**

a. Blood glucose concentration: continuous
b. Number of children: ordinal
c. Job classification: nominal
d. All of the above are incorrect
e. None of the above are incorrect

Q3: **True or False:** When we fail to reject the null hypothesis based on our data, we have proven that the alternative is false and the null is true.

Q4: **True or False:** Alpha (α) should always be set at 0.05.

Answers to Self Evaluation:

Q1: **False**. In statistical testing, we are actually testing the null hypothesis.

Q2: **E.** All of the types of data are correctly classified.

Q3: **False.** When we fail to reject the null hypothesis, we are concluding that our data does not support that the null is true. What we are concluding is that there is not enough evidence to support rejection of the null as false. This can be due to a number of things such as small sample size, or just a rare sample.

Q4: **False.** Alpha of 0.05 is an arbitrary value set by the researcher. There may be situations where a researcher is comfortable with having a lower probability of failing to reject the null hypothesis when the null is "true".

ERIC NOTEBOOK IS PRODUCED BY THE EDUCATIONAL ARM (MICHEL A. IBRAHIM, MD, PHD, DIRECTOR) OF THE EPIDEMIOLOGIC RESEARCH AND INFORMATION CENTER AT DURHAM, NC (RON HORNER, PHD, DIRECTOR)

If you are not currently receiving ERIC Notebook and would like to be added to the distribution list, please fill out the form below:

Name: _____

Degree(s):_____

Address:_____

City, State, Zip:_____

Telephone Number:_____

Fax Number:_____

E-mail Address:_____

Affiliation: VA_____ Other Gov't_____
Academic _____ Private_____
Please fax to: 919-416-5836 – Attn: Beth Armstrong **Or**
Mail to: Beth Armstrong, ERIC Program Manager, VA Medical Center (152), 508 Fulton Street, Durham, NC 27705

Upcoming Topics

- **Causality**
- **Healthcare Epidemiology**

Please let Beth Armstrong know which topics are of special interest to you so that we can include them in a future issue.

Reminder:

ERIC Notebook can be viewed online at

http://sph.unc.edu/courses/eric or

http://hsrd.durham.med.va.gov/ERI

BETH ARMSTRONG, ERIC PROGRAM MANAGER
VA MEDICAL CENTER (152)
508 FULTON STREET
DURHAM, NC 27705